基于注意力机制的卷积神经网络与数据增强提升白血病细胞分类

Douglas Costa Braga 和 Daniel Oliveira Dantas塞尔希培联邦大
学计算机系，巴西塞尔希培州圣克里斯托旺市¸˜，˜，´，˜
contato@douglasbraga.com，ddantas@dcomp.ufs.br

arXiv:2601.01026v1 [cs.CV] 3 Jan 2026

关键词: 白血病分类、急性淋巴细胞白血病、卷积神经网络、注意力机制、压缩与激励、EfficientNetV2、细胞分类、医学图像分析、计算机辅助诊断、可复现机器学习、深度学习流程、实验稳健性

摘要: 我们提出了一套可复现的白血病细胞分类深度学习流程，重点关注系统架构、实验稳健性以及医学图像分析中的软件设计选择。急性淋巴细胞白血病（ALL）是最常见的儿童癌症，需要专家进行显微镜诊断，但诊断结果会受到观察者间差异和时间限制的影响。所提出的系统将基于注意力机制的卷积神经网络与 Squeeze-and-Excitation 机制相结合，并采用 EfficientNetV2-B3，实现 ALL 细胞自动分类。我们采用全面的数据增强、用于应对类别不平衡的焦点损失，以及按患者划分数据，以确保评估稳健且可复现。在 C-NMC 2019 数据集（包含来自 62 名患者的 12,528 张原始图像）上，该系统在测试集上取得了 97.89% 的 F1分数和 97.89% 的准确率；通过进行 100 次迭代的蒙特卡洛实验进行统计验证，证实其相较于基线方法有显著提升（$p < 0.001$）。所提出的流程比现有方法最多提升 4.67%，同时参数量比 VGG16 少 89%（15.2M 对比 138M）。注意力机制能够生成诊断相关细胞特征的可解释可视化结果，表明现代基于注意力机制的架构能够提升白血病细胞分类性能，同时保持适合临床部署的计算效率。

1 引言

急性淋巴细胞白血病（ALL）的特征是骨髓中未成熟淋巴母细胞过度增殖，是最常见的儿童癌症，发病高峰年龄为2至5岁（Pui等，2012）。

目前，白血病诊断的金标准是检查骨髓穿刺液。然而，这是一种侵入性操作，有时可能更倾向于检查外周血，尽管其准确性较低（Metrock等，2017）。此外，外周血检查过程耗时费力，需要经过培训的人员操作，且观察者间差异很大（Park等，2024）。这种主观性，加上资源受限环境中专业人员有限，使得客观、自动化诊断工具的需求尤为迫切。

计算机辅助诊断（CADx）系统应运而生，旨在通过提供客观、自动化的显微血细胞图像分析来克服这些局限。早期方法依赖手工特征提取，采用全面的特征集取得了显著成效。MoradiAmin（MoradiAmin 等，2016年）结合纹理、形状和颜色描述符，准确率达到 96.37%；Sant’Anna（Sant’Anna 等，2022年）则使用统计、形态学和纹理特征及集成分类器，报告 F1分数为 93.70%。

随着卷积神经网络的发展，研究范式转向了深度学习。卷积神经网络通过自动特征学习展现出更优越的性能（Sampathila 等人，2022；Talaat 和 Gamel，2023）。使用预训练网络的迁移学习方法显示出良好前景，其中基于 AlexNet 的方法达到了超过 97% 的准确率（Shafique 和 Tehsin，2018；Rehman 等人，2018）。近期研究探索了包括 VGG 变体（Oliveira 和 Dantas，2021）、ResNet（Pan 等，2019）以及 Xception网络在内的先进架构，用于恶性细胞分类。

过去三年中，该领域取得了显著进展。近期研究探索了将视觉Transformer作为传统卷积神经网络的替代方案（Oybek Kizi 等人，2025），与此同时，注意力机制在多种架构中日益受到重视（Jawahar 等人，2024；Gokulkannan 等人，2024）。这些进展进一步凸显了高效注意力机制方法对于临床部署的重要性。另一方面，使用深度学习模型时，可解释性仍然是一个值得关注的问题。Abhishek（Abhishek 等人，2023）使用梯度加权类激活映射（Grad-CAM）来可视化图像中的相关特征。

当前方法存在严重局限，阻碍了其在临床上的应用：(1) 计算复杂度高，需要大量资源（VGG16：1.38 亿个参数），限制了临床部署；(2) 缺乏可解释性，犹如“黑箱”，诊断透明度不足；(3) 对数据集不平衡的处理不足；(4) 评估协议不一致，无法对不同研究进行公平比较。

本文通过一种新颖的基于注意力的卷积神经网络架构解决上述局限，该架构结合了压缩与激励机制（Hu 等人，2018）和 EfficientNetV2-B3 骨干网络（Tan 和 Le，2021）。我们的方法采用焦点损失（Lin 等人，2017）处理类别不平衡，并按照 Mourya（Mourya 等人，2018）的建议，按患者划分数据，以确保对泛化能力进行稳健评估。

我们的贡献包括：(1) 提出一种高效架构，以比 VGG16 少 89% 的参数实现最先进性能（F1 分数为 97.89%）；(2) 提供可解释的注意力可视化，突出显示与诊断相关的区域；(3) 通过全面的数据增强处理数据集不平衡；(4) 进行严格评估，证明该方法在 C-NMC 2019数据集上取得了具有统计显著性的改进。

从软件工程的角度来看，本研究强调可复现性、模块化流水线设计以及统计上稳健的评估协议，这些都是在临床环境中部署深度学习系统的关键要求。

2 方法

图 1 展示了我们用于恶性细胞分类的方法框架，该框架将 EfficientNetV2-B3 与压缩与激励注意力机制相结合。该方法的主要步骤包括数据预处理、数据增强、评估和验证集。所提出的实现

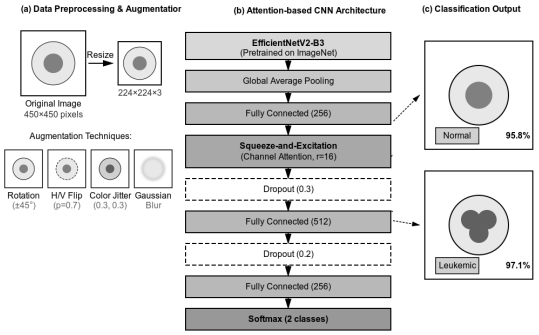


图1：所提出方法的概览：(a) 数据预处理与数据增强，(b) 基于注意力的卷积神经网络架构，以及 (c) 分类结果。

可在 GitHub 上获取。[1]

2.1 数据集

我们使用了由 SBILab 研究团队创建的 C-NMC 2019数据集（SBILab，2022）。该数据集作为 ISBI 2019 举办的*B-ALL*白细胞癌显微图像中正常细胞与恶性细胞的分类挑战赛的一部分发布。数据集包含来自B细胞急性淋巴细胞白血病（B-ALL）患者的淋巴母细胞显微图像，以及来自健康个体的正常淋巴细胞图像。

该数据集分为三个文件夹：训练数据、初步测试数据和最终测试数据。训练数据包含来自73名受试者的10,661张图像。初步测试数据包含来自28名受试者的1,867张图像。最终测试数据包含来自17名受试者的2,586个未标注细胞，我们未在实验中使用这些数据。

所有图像均由 SBI- Lab 团队进行了预处理，包括分割、增强和染色归一化（Duggal 等，2016；Gupta 等，2017；Duggal 等，2017）。每张图像包含 450×450像素，中央位置有一个经过分割的淋巴细胞，背景为黑色。细胞采用 Jenner–Giemsa 技术染色。

数据集按受试者层级进行整理，以确保评估得当且不受受试者特异性偏差影响。在实验中，我们合并了训练集和初步测试数据，得到来自 101 名患者的 12,528 张带标签图像（60 名急性淋巴细胞白血病患者，共 8,491 张图像；41 名血液学患者，共 4,037 张图像）。

我们采用**按患者划分**，以确保进行稳健的泛化评估。这101名患者被

[1]可通过 https://github.com/*****/***** 获取

表 1：数据集划分中患者和图像的分布

类别	总计 患者	训练集 患者（图像）	验证集 患者（图像）	Test 患者（图像）
ALL（恶性）	60	48 (7,324)	6 (616)	6 (1,102)
HEM（健康）	41	32 (7,324*)	4 (983)	5 (982)
总计	**101**	**80（14,648）**	**10（1,599）**	**11（2,084）**

* 为实现类别平衡，将 HEM 增强至 7,324 个样本。

分为训练集（80名患者）、验证集（10名患者）和测试集（11名患者），确保所有划分中的类别分布具有代表性。

各数据集划分中的患者分布：

- **训练集**：共 80 名患者（48 例 ALL + 32HEM）；
- **验证集**：共 10 位患者（6 位 ALL + 4HEM）；
- **测试集**：共 11 位患者（6 位 ALL + 5HEM）。

表 1 展示了各数据集划分中患者和图像的最终分布。训练集存在类别不平衡（原始图像中 ALL/HEM 比例为 2.89），我们仅对少数类（HEM）进行全面的数据增强来解决这一问题。数据增强在训练过程中动态执行，使 HEM 训练集的图像数量从 2,533 张增加到 7,324 张，从而实现类别平衡。验证集和测试集仅保留原始图像，不进行任何数据增强，以确保评估结果公正无偏。

严格遵循按患者划分的方案，确保我们报告的性能指标能够反映模型对未见患者的真实泛化能力，这对临床部署至关重要。

2.2 预处理

训练前，我们进行了预处理以提升图像质量和一致性。我们使用双线性插值将每张图像从原始的 450×450 像素调整为 384×384 像素，以保持细胞结构的平滑性。我们使用 ImageNet 数据集的数值（均值 = [0.485、0.456、0.406], ；标准差 = [0.229、0.224、0.225]）进行按通道归一化，使数据分布与预训练数据集保持一致。

2.3 数据增强

数据增强可应对类别不平衡、增强模型泛化能力并缓解过拟合。我们使用 PyTorch tr ansforms 实现了保守的数据增强流程：

- **几何变换**：随机水平翻转（$p = 0.5$）和旋转（$\pm 10°$）。
- **颜色变换**：随机调整亮度和对比度（0.1，0.1），以保留

对准确分类至关重要的细微细胞特征。

训练期间，仅对少数类（HEM）通过在线变换执行这些操作。数据增强将 HEM 训练集从 2,533 张原始图像扩充至 7,324 张增强图像，从而与 ALL 类别（7,324 张图像）实现完全平衡。增强后的训练集包含 14,648 张图像（7,324 张 ALL + 7，324 张增强后的 HEM 图像）。对于验证集和测试集，我们仅执行确定性预处理步骤（将图像调整为 384×384 像素并进行归一化），以确保评估一致且可复现。焦点损失与保守的数据增强相结合，有助于缓解类别不平衡的影响，同时保留白血病诊断所必需的细胞特征。

2.4 模型训练

我们的模型架构能够捕捉恶性细胞的局部和全局特征，同时通过注意力机制提供可解释性。

2.4.1 骨干网络

我们选择 EfficientNetV2-B3（tf efficientnetv2 b3）作为骨干网络，因为它兼顾了性能与计算效率。EfficientNetV2 通过引入融合 MBConv 模块并优化网络缩放，对原始 EfficientNet 进行了改进；与类似架构相比，它以更少的参数实现了更高的准确率（Tan 和 Le，2021）。我们使用 ImageNet 预训练权重初始化骨干网络，以利用迁移学习的优势，并将其配置为 num classes=0和 global pool=”，以直接提取特征图。我们通过丢弃 _（比率 = 0.3）和随机深度（drop path 比率 = 0.2）对骨干网络进行正则化，以防止过拟合。

2.4.2 注意力机制

我们在全局池化之前，将压缩与激励（SE）注意力机制（Hu 等人，2018）应用于特征图。该机制首先通过全局平均池化将空间信息压缩为通道描述符，并据此自适应地重新校准通道级特征响应。随后，瓶颈结构通过两个全连接层对通道间的相互依赖关系进行建模：第一个全连接层将维度缩小 16 倍（压缩比），并接入 ReLU 激活；第二个全连接层则恢复原始通道维度，并接入 Sigmoid 激活，以生成注意力

权重，取值范围为 [0,1]。这些权重与原始特征图逐元素相乘，以突出与诊断相关的模式，同时抑制无关模式，从而提升分类性能和模型可解释性。

2.4.3 分类头

在 SE注意力模块之后，应用全局平均池化以获得固定大小的特征向量，随后通过多层类别分类头进行处理。该架构包括：丢弃（比率 = 0.3）；一个包含 512 个神经元的全连接（FC）层，后接批归一化和 ReLU 激活；丢弃（比率 = 0.2）；一个包含 256 个神经元的 FC 层，后接批归一化和 ReLU 激活；以及一个最终的 FC 层，该层包含两个输出神经元，对应于二分类任务（健康与恶性）。在训练集期间，焦点损失会在内部对这些 Logits 应用 Softmax，而在推理期间则使用显式 Softmax 来获得类别概率。

2.4.4 损失函数

我们实现了焦点损失，以解决数据集中固有的类别不平衡问题。焦点损失会降低分类正确样本的贡献，并将学习重点放在困难的、分类错误的样本上。对于二分类，焦点损失定义如下：

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \qquad (1)$$

其中，p_t =exp(−CE)是模型对正确类别估计的概率，基于交叉熵损失CE得出；α_t 是类别专属的平衡因子（类别1为α，类别0为 $1-\alpha$），γ是聚焦参数，用于控制降低简单样本权重的幅度。我们根据经验验证设定了 $\alpha = 0.25$ 和$\gamma = 2.0$，其中较低的α 值可提高少数类健康类别的权重，以弥补类别不平衡。

2.4.5 训练策略

我们采用AdamW优化器，初始学习率为 1×10^{-4}，权重衰减为 1×10^{-5}，beta值为 (0.9,0.999)。AdamW将权重衰减与梯度更新解耦，从而促进更好的泛化（Loshchilov和Hutter，2019）。我们采用OneCycleLR调度器进行高效训练，最大学习率设为 1×10^{-3}，并使用余弦退火调度策略。

由于数据集存在类别不平衡，我们采用了焦点损失（Lin等人，2017），并以 $\alpha = 0.25$和 $\gamma = 2.0$ 作为损失函数，着重于从难以分类的样本中学习，并降低分类正确样本的贡献。此外，我们采用了梯度裁剪，将最大范数设为1.0，以稳定训练集。

我们采用批量大小为 8 的设置，最多训练 100 个训练轮次，并实施早停以防止过拟合。如果验证集 F1 分数连续 10 个训练轮次至少未提升 0.002，则终止训练。模型选择标准优先考虑验证集 F1 分数高于 0.85，同时尽量缩小训练集与验证集 F1 分数之间的差距，以确保泛化能力。我们使用 PyTorch 的自动混合精度（AMP）进行混合精度训练，以优化内存使用和计算效率。

训练期间的数据增强包括随机水平翻转（概率为 0.5）、随机旋转（$\pm 10^{\circ}$）以及颜色抖动（亮度和对比度变化为 ± 0.1）。所有图像均调整为 384×384 像素，并使用 ImageNet 统计数据进行归一化。

实验在配备 Intel Core i7 处理器、32GB RAM 和 NVIDIA GeForce RTX 4060 GPU（8GB 显存）的系统上进行。模型使用支持 CUDA 的 PyTorch 1.9.0 实现。

2.5 评估指标

我们使用多种指标评估模型性能：准确率、精确率、召回率、F1分数和 ROC 曲线下面积（AUC）。由于数据集类别不平衡（HEM样本数量多于 ALL 样本），我们更加重视 F1分数和 AUC，因为与单独使用准确率相比，这些指标能够提供更均衡的评估。F1分数是精确率和召回率的调和平均值，尤其适用于类别不平衡的二分类任务。

2.6 统计验证

为确保结果的鲁棒性和统计显著性，我们参照 Sant'Anna（Sant'Anna 等，2022）的方法进行了包含 100 次迭代的蒙特卡洛实验。每次迭代中，我们随机将患者重新分配至训练集、验证集和测试集，同时保持各集合原有比例，分别约为 79%、15% 和 6%。这种方法可评估模型在不同患者组合下的鲁棒性，相较于单次固定划分，能够提供更全面的评估。

在统计验证中，我们将基于注意力机制的 EfficientNetV2-B3模型与一个架构相同但不含 Squeeze-and-

表 2：所提出模型在测试集（2,084 张图像）上的混淆矩阵。

预测标签		HEM	ALL
真实标签	HEM	964	18
	ALL	26	1,076

激励（SE）注意力模块的模型进行了比较，同时保持所有其他超参数不变。

2.7 实现细节

我们的实现采用 PyTorch 1.9.0 和 CUDA 11.1。SE 注意力模块集成在全局平均池化层之前，压缩比为 16。分类器头由三个全连接层组成，维度分别为 512、256 和 2（输出类别数），并结合批归一化和随机失活正则化（比率分别为 0.3、0.2 和 0.3）以防止过拟合。骨干网络的丢弃率为 0.3，随机深度率为 0.2。

报告的结果是在训练期间验证集上取得的最佳性能，根据验证集 F1 分数超过 0.85 且训练–验证差距最小的标准选出。模型开发过程中未使用测试集标签。蒙特卡洛验证采用分层抽样，以确保统计严谨性，同时在各次迭代中保持类别分布。

3 结果与讨论

3.1 所提模型的性能

我们提出的基于注意力机制的卷积神经网络模型采用 EfficientNetV2-B3 架构，在区分健康（HEM）细胞与恶性（ALL）细胞方面表现出卓越的效果，准确率、精确率、召回率和 F1分数均达到 97.89%。该模型出色的区分能力还体现在其 93.77% 的 AUC 上，表明模型在不同分类阈值下均具有稳健的性能。

混淆矩阵（表 2）显示，模型在测试集上正确分类了 964 个健康细胞和 1,076 个恶性细胞，同时将 18 个健康细胞误分类为恶性细胞，将 26 个恶性细胞误分类为健康细胞。这表明模型对恶性细胞具有较高的敏感度（97.6%），对健康细胞具有较高的特异度（98.2%），说明其对这两类细胞的分类效果都非常出色。

图 2 展示了准确率、损失和 F1 分数的训练和验证曲线。模型平稳收敛，并在第 10 个训练轮次达到最高验证集 F1 分数 98.37%。训练持续至第 18 个训练轮次触发早停，并保留第 10 个训练轮次的模型作为最终模型，以防止过拟合。

3.2 比较分析

我们将本模型与传统特征提取方法以及最先进的卷积神经网络架构进行了比较。表 3 展示了不同方法在 C-NMC 2019数据集上的 F1分数。

在 C-NMC 2019数据集上评估的各方法中，本模型取得了最高的 F1分数（97.89%）。与 Sampathila（Sampathila 等人，2022）报告的 95.43% 相比，这一结果显著提高了 2.46 个百分点。

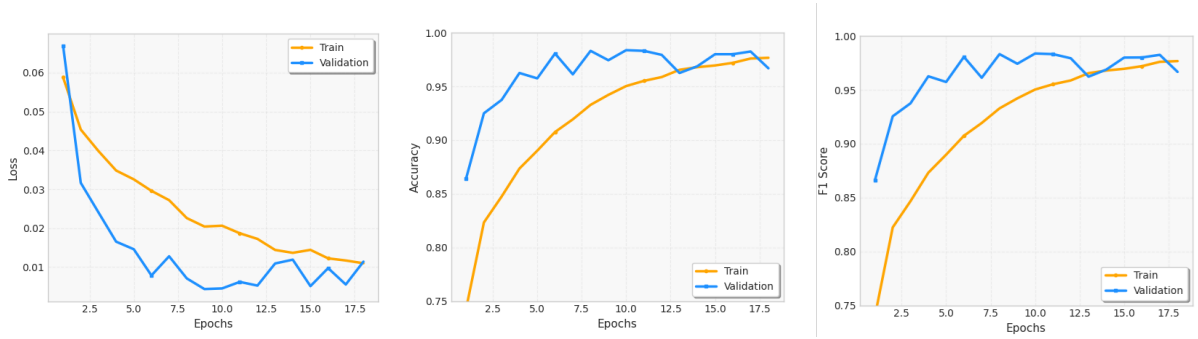
所提出的架构需要 1,520 万个参数，与 VGG16（1.384 亿）和 VGG19（1.437 亿）相比，参数量减少了约 89%。这一效率与近期的轻量级架构（如 Sampathila 的 ALLNET（Sampathila 等人，2022））相当，同时实现了显著更高的 F1分数（97.89% 对 95.43%）。EfficientNetV2-B3 骨干网络与 SE注意力模块的结合，实现了性能与计算效率之间的平衡。

这些结果表明，注意力机制能够在不成比例增加模型复杂度的情况下，显著提升分类性能，这对于资源受限的临床环境尤其有益。

3.3 统计显著性分析

我们使用按患者随机化的方法进行了 100 次迭代的蒙特卡洛实验，以评估模型的鲁棒性。所提出的方法表现出卓越的稳定性，平均 F1分数为 98.15% ± 0.41%（95% 置信区间： [97.11%，98.72%]），AUC 为 99.80% ± 0.07%（95% 置信区间： [99.62%，99.91%]）。所有分类指标（准确率、精确率、召回率）均表现出类似的鲁棒性，平均值为 98.16% ± 0.41%。较窄的置信区间——分类指标的跨度小于 1.6 个百分点，AUC 的跨度为 0.3 个百分点——证实了我们的方法在不同患者分布下的鲁棒性。

蒙特卡罗实验的平均F1分数（98.15 ± 0.41%）比单一测试集结果（97.89%）高出0.26个百分点。这种差异在以下几方面均属预期且具有统计一致性：



(a) Loss (b) 准确率 y (c) F1分数

图 2：训练集和验证集曲线显示了 (a) 损失、(b) 准确率和 (c) F1分数的变化情况。在第 10 个训练轮次时，验证集性能达到最佳，F1分数为 98.37%。

表 3： 各方法在 C-NMC 2019 数据集

方法	F1分数	方法
提出的模型	**97.89%**	**EfficientNetV2-B3 +SE注意力**
(Sampathila 等人，2022)	95.43%	采用数据增强的定制卷积神经网络
(Talaat 和 Gamel，2023)	94.07%	采用超参数优化的卷积神经网络
(Sant'Anna 等，2022年)	93.70%	特征提取 + 集成（人工神经网络+支持向量机+朴素贝叶斯）
(Oliveira 和 Dantas，2021年)	92.60%	使用数据增强的 VGG16
(Pan 等，2019年)	92.50%	迁移学习 残差网络 +校正
(Honnalgere 和 Nayak，2019年)	91.70%	迁移学习 VGG16
（Xiao等，2019年）	90.30%	多模型集成
（Verma和Singh，2019年）	89.47%	迁移学习 MobileNetV2
(Prellberg和Kramer，2019年)	87.89%	从头训练的ResNeXt50
(Shah等，2019年)	87.58%	迁移学习卷积神经网络-RNN
(Marzahl 等人，2019)	87.46%	迁移学习 ResNet18
(Ding 等人，2019)	86.74%	InceptionV3、DenseNet、InceptionResNetV2
（Kulhalli 等人，2019）	85.70%	ResNeXt50 和 ResNeXt101
（Liu 和 Long，2019）	84.00%	迁移学习 Inception + 残差网络
（Khan 和 Choo，2019）	81.79%	迁移学习 残差网络 + 挤压激励网络

(1) 蒙特卡罗方法对100种不同的患者组合进行性能平均，降低了特别棘手病例的影响；(2) 仅包含11名患者的固定测试集可能恰好包含更多难以分类的样本；(3) 单一测试集结果（97.89%）完全落在蒙特卡罗方法的95%置信区间[97.11%、98.72%], 内，证实了两种评估方法在统计上的一致性。这种双重验证策略——使用固定测试集与其他研究直接比较，并使用蒙特卡罗方法进行稳健的统计验证——为模型在不同患者群体中的泛化能力提供了全面证据。

3.4 消融研究

我们开展了消融研究，以系统评估每个组件的贡献。

表 4：验证集上的消融研究结果

配置	F1分数	Δ
完整模型	**97.89%**	–
不进行数据增强	93.50%	-3.77%
不使用注意力机制	94.93%	-2.34%
不使用焦点损失	95.84%	-1.43%

表 4 展示了其对模型性能的定量影响。

消融实验结果表明，数据增强影响最大（3.77 个百分点），验证了其在应对训练数据有限和提升模型泛化能力方面的重要性。SE注意力机制带来 2.34 个百分点的提升，同时通过特征重新校准提供可解释性，相关讨论见第 3.5 节。焦点损失通过更好地处理类别

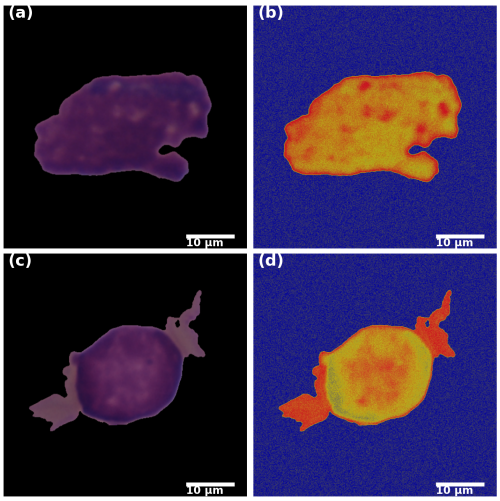


图3：注意力图可视化：(a、c) 恶性细胞和健康细胞的原始图像，(b、d) 对应的注意力图，突出显示诊断相关区域。

不平衡，将分类正确的样本降权并着重处理难以分类的样本，使性能提升了 1.43 个百分点。这些组件的累积效果表明，它们协同作用，共同实现了最先进的性能。

3.5 注意力可视化

我们对压缩与激励模块生成的注意力图进行了可视化。图3展示了恶性细胞和健康细胞的示例及其对应的注意力图。

注意力图展示了模型对各类别中临床相关特征的关注：

- 恶性细胞：模型重点关注不规则的细胞核形态、升高的核质比以及异常的染色质模式——这些都是公认的诊断标准（Bennett等人，1976年）。
- 健康细胞：注意力图突出显示规则的细胞轮廓和均匀的染色质分布。

4 结论

我们提出了一种新颖的自动化白血病细胞分类方法，将 EfficientNetV2-B3 与压缩与激励注意力机制相结合。该方法在 C-NMC 2019数据集上取得了最先进的性能，在测试集上的 F1分数达到 97.89%，同时所需参数比基于 VGG16 的方法少 89%。经过 100 次迭代的蒙特卡洛验证结果表明，该方法具有出色的鲁棒性（F1分数：98.15 ± 0.41%），证实其在不同患者分布下具有很强的泛化能力。

我们方法的一项关键优势在于注意力机制所带来的可解释性。注意力图可视化显示，模型关注的是与临床相关的细胞特征——不规则的细胞核形态、较高的核质比以及染色质模式——从而揭示其决策过程。这种可解释性有助于建立对人工智能辅助诊断系统的信任，并使临床医生能够理解和验证模型预测，从而促进其临床应用。

我们全面的数据增强流程结合焦点损失和按患者划分数据，有效应对了数据集不平衡和样本量有限的挑战。消融研究量化了各项措施的独立贡献：数据增强（3.77%）、SE注意力（2.34%）和焦点损失（1.43%），证实了它们对分类性能的协同作用。这种方法在医学影像应用中尤为重要，因为受专家标注需求和患者隐私问题的限制，难以获取规模庞大且带有标注的数据集。

我们在所有实验中严格采用按患者划分方案，确保所报告的性能指标能够反映模型对未见患者的真实泛化能力，这是临床部署的一项关键要求。数据泄漏几乎为零，且在不同患者分配方案下性能始终稳定，这验证了我们方法的临床适用性。

我们的研究存在一些值得考虑的局限性。首先，在来自不同机构、采用不同染色方案和成像设备且患者人口统计特征各异的外部数据集上进行验证，将有助于进一步评估模型在多样化临床环境中的泛化性。其次，尽管 SE注意力提供了一定的可解释性，但采用更先进的可解释性技术，例如反事实解释或基于概念的可解释性，或可进一步增强临床信任并促进错误分析。

未来的研究方向包括：(1) 将我们的 SE注意力机制 与视觉 Transformer 相结合，在保持计算效率的同时兼顾局部特征强调与全局上下文建模；(2) 开展跨数据集训练和评估，以提升模型在不同成像条件下的鲁棒性（Oybek Kizi 等，2025）；(3) 在 C-NMC 2019 无标注测试集（2,586 张图像）上应用自监督预训练，以便在监督微调前充分利用额外数据（Kazeminia 等，2024）；(4) 将我们的注意力机

制扩展至外周血涂片图像与流式细胞术数据的多模态融合（Cheng 等，2024），并采用跨模态注意力；以及 (5) 开展前瞻性临床验证研究，评估其在真实世界中的性能及其融入诊断工作流程的可行性。

致谢

作者感谢使用 Claude（Anthropic，2024）作为 AI 助手，协助完成手稿准备工作，包括润色文本、审阅技术写作以及检查语法和拼写。所有实验设计、实施、数据分析、结果解读和科学结论均为作者原创，并由作者全权负责。

参考文献

Abhishek, A., Jha, R., Sinha, R., 和 Jha, K.（2023）。利用深度迁移学习，在主体无关测试数据集上自动检测和分类白血病，并辅以 Grad-CAM 可视化。《生物医学信号处理与控制》，83：104722。

Anthropic（2024）。Claude Sonnet 4.5。https://www.anthropic.com/claude。大语言模型，访问日期：2025-11-20。

Bennett, J. M., Catovsky, D., Daniel, M.-T., Flandrin, G., Galton, D. A., Gralnick, H. R., 和 Sultan, C.（1976）。急性白血病分类提案。法美英（FAB）协作组。《英国血液学杂志》，33(4)：451–458。

Cheng, F.-M., Lo, S.-C., Lin, C.-C., Lo, W.-J., Chien, S.- Y., Sun, T.-H., 和 Hsu, K.-C.（2024）。利用深度学习通过急性白血病定向管进行流式细胞术，辅助急性白血病检测和细胞分类。《科学报告》，14:8350。

Ding, Y., Yang, Y., 和 Cui, Y.（2019）。用于白细胞癌分类的深度学习。载于《生物工程讲义》，第33–41页。新加坡施普林格。

Duggal, R., Gupta, A., Gupta, R., 和 Mallick, P.（2017）。SD-Layer：用于医学显微成像中卷积神经网络的染色解卷积层。载于国际医学图像计算与计算机辅助干预会议，第435–443页。施普林格。

Duggal, R., Gupta, A., Gupta, R., Wadhwa, M., and Ahuja, C. (2016). 使用深度置信网络对显微图像中的重叠细胞核进行分割。载于 第十届印度计算机视觉、图形学与图像处理会议论文集，第 1–8 页。美国计算机协会。

Gokulkannan, K., Raj, S., Kumar, M., and Prasad, A. (2024). 用于高效白血病检测的多尺度自适应与注意力机制扩张卷积神经网络模型，以及基于多尺度 trans-res-Unet3+ 的分割网络。生物医学信号处理与控制, 90:105847。

Gupta, R., Mallick, P., Duggal, R., Gupta, A., 和 Sharma, O.（2017）。多发性骨髓瘤中浆细胞显微图像的染色颜色归一化与分割：开发计算机辅助自动疾病诊断工具的前奏。载于《临床淋巴瘤、骨髓瘤与白血病》，第17卷，页码e99。爱思唯尔。

Honnalgere, A. 和 Nayak, G. (2019). B-ALL 白细胞癌显微图像中正常细胞与恶性细胞的分类。载于《生物工程讲义》，第 1–12 页。新加坡施普林格。

Hu, J., Shen, L. 和 Sun, G. (2018). Squeeze-and-excitation 网络。载于*IEEE* 计算机视觉与模式识别会议（*CVPR*）论文集，第 7132–7141 页。

Jawahar, M., Anbarasi, L., Narayanan, S., Gowri, S. 和 Kumar, S. (2024). 一种基于注意力机制的急性淋巴细胞白血病分类深度学习方法。《科学报告》，14:17447。

Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N. 和 Albarqouni, S. (2024). 用于急性髓系白血病分类的自监督多实例学习。*arXiv*预印本*arXiv:2403.05379*。

Khan, M. A. 和 Choo, J.（2019）。通过卷积神经网络对癌症显微图像进行分类。载于《生物工程讲义》，第 141–147 页。新加坡施普林格。

Kulhalli, R., Savadikar, C. 和 Garware, B.（2019）。迈向 B细胞急性淋巴细胞白血病的自动分类。载于《生物工程讲义》，第 63–72 页。新加坡施普林格。

Lin, T.-Y., Goyal, P., Girshick, R., He, K. 和 Doll´ar, P.（2017）。用于密集目标检测的焦点损失。载于*IEEE*国际计算机视觉会议（*ICCV*）论文集，第 2980–2988 页。

Liu, Y. 和 Long, F.（2019）。基于深度装袋集成学习的急性淋巴细胞白血病细胞图像分析。载于《生物工程讲义》，第 113–121 页。新加坡施普林格。

Loshchilov，I. 和 Hutter，F.（2019）。解耦权重衰减正则化。载于国际学习表征会议。

Marzahl, C., Aubreville, Voigt, M., J¨orn, 和 Maier, A.（2019）。基于注意力机制的深度学习方法与高级数据增强技术对血涂片显微图像中白血病性B淋巴母细胞的分类。载于《生物工程讲义》，第13–22页。新加坡施普林格。

Metrock, L. K., Summers, R. J., Park, S., Gillespie, S., Castellino, S., Lew, G., 和 Keller, F. G.（2017）。流式细胞术进行外周血免疫表型分析在儿童急性白血病诊断中的应用价值。《儿科学、血液与癌症》，64(10)：e26526。

MoradiAmin, M., Memari, A., Samadzadehaghdam, N., Kermani, S., 和 Talebi, A.（2016）。基于显微图像分析的急性淋巴细胞白血病细胞亚型计算机辅助检测与分类。《显微镜研究与技术》，79(10):908–916。

Mourya, S., Kant, S., Kumar, P., Gupta, A., 和 Gupta, R.（2018）。LeukoNet：用于分类 B细胞急性淋巴细胞白血病中正常细胞与白血病原始细胞的基于DCT的卷积神经网络架构。*arXiv* 预印本 *arXiv:1810.07961*。

Oliveira, J. E. M. d. 和 Dantas, D. O.（2021）。结合数据增强和卷积神经网络对正常细胞与白血病细胞进行分类。载于第*16*届国际计算机视觉、成像与计算机图形理论联合会议论文集，第4卷，VISAPP，第685–692页。INSTICC，SciTePress。

Oybek Kizi, R. F., Theodore Armand, T. P., 和 Kim, H.- C. (2025). 基于血涂片图像的白血病分类深度学习技术综述。《style id='1'>AppliedBiosciences，4(1):9.

Pan, Y., Liu, M., Xia, Y., 和 Shen, D. (2019). 用于正常细胞和恶性细胞分类的邻域校正算法。载于生物工程讲义，第 73–82 页。新加坡施普林格。

Park, S., Park, Y. H., Huh, J., Baik, S. M., 和 Park, D. J. (2024). 通过髓母细胞和淋巴母细胞分类，利用外周血细胞图像区分急性髓系白血病与急性淋巴细胞白血病的深度学习模型。《 style id='1'>《数字健康》，10:20552076241258079.

Prellberg, J. 和 Kramer, O. (2019). 使用卷积神经网络对显微图像中的急性淋巴细胞白血病进行分类。载于生物工程讲义，第 53–61 页。新加坡施普林格。

Pui, C.-H., Yang, J. J. 和 Hunger, S. P.（2012）。儿童急性淋巴细胞白血病：我们将走向何方，又该如何实现？《血液》，120(6)：1165–1174。

Rehman, A., Abbas, N., Saba, T., Rahman, S. U., Mehmood, Z. 和 Kolivand, H.（2018）。使用深度学习对急性淋巴细胞白血病进行分类。显微镜研究与技术，81(11)：1310–1317。

Sampathila, N., Chadaga, K., Goswami, N., Chadaga, R. P., Pandya, M., Prabhu, S., Bairy, M. G., Katta, S. S., Bhat, D. 和 Upadya, S. P.（2022）。基于血涂片图像检测急性淋巴细胞白血病的定制化深度学习分类器。《*Healthcare*》，10(10)：1812。

Sant’Anna, Y. F. D. d.、Oliveira, J. E. M. d. 和 Dantas, D. O.（2022）。正常细胞与白血病细胞的可解释轻量级集成分类。《Computers》，11(8)：125。

SBILab（2022）。信号处理与生物医学成像实验室。在线获取：http://sbilab.iiitd.edu.in/（访问日期：2022年6月8日）。Shafique, S. 和 Tehsin, S.（2018）。使用预训练深度卷积神经网络检测急性淋巴细胞白血病并对其亚型进行分类。《癌症研究与治疗技术》，17：1533033818802789。Shah, S.、Nawaz, W.、Jalil, B. 和 Khan, H. A.（2019）。结合卷积神经网络和循环神经网络对B-ALL癌症中的正常细胞和白血病原始细胞进行分类。载于《生物工程讲义》，第23–31页。新加坡施普林格。Talaat, F. M. 和 Gamel, S. A.（2023）。使用C-NMC白血病数据集进行白血病检测与分类中的机器学习。《多媒体工具与应用》，83(3)：8063–8076。Tan, M. 和 Le, Q. V.（2021）。EfficientNetV2：更小的模型与更快的训练。载于《第38届国际机器学习会议论文集》，《机器学习研究论文集》第139卷，第10096–10106页。PMLR。Verma, E. 和 Singh, V.（2019）。2019年ISBI挑战赛：用于B-ALL细胞分类的卷积神经网络。载于《生物工程讲义》，第131–139页。新加坡施普林格。Xiao, F.、Kuang, R.、Ou, Z. 和 Xiong, B.（2019）。Deep-MEN：用于B淋巴母细胞分类的多模型集成网络。载于《生物工程讲义》，第83–93页。新加坡施普林格。